

# ZINCPharmer: pharmacophore search of the ZINC database

David Ryan Koes\* and Carlos J. Camacho

Department of Computational and Systems Biology, University of Pittsburgh, 3501 Fifth Avenue, Pittsburgh, PA 15260, USA

Received January 16, 2012; Revised April 4, 2012; Accepted April 12, 2012

## ABSTRACT

**ZINCPharmer (<http://zincpharmer.csb.pitt.edu>) is an online interface for searching the purchasable compounds of the ZINC database using the Pharmer pharmacophore search technology. A pharmacophore describes the spatial arrangement of the essential features of an interaction. Compounds that match a well-defined pharmacophore serve as potential lead compounds for drug discovery. ZINCPharmer provides tools for constructing and refining pharmacophore hypotheses directly from molecular structure. A search of 176 million conformers of 18.3 million compounds typically takes less than a minute. The results can be immediately viewed, or the aligned structures may be downloaded for off-line analysis. ZINCPharmer enables the rapid and interactive search of purchasable chemical space.**

## INTRODUCTION

A pharmacophore describes the structural arrangement of the essential molecular features of an interaction between a ligand and its receptor. Searching chemical libraries for compounds that match a specific pharmacophore is an established method of virtual screening (1–3). The two main challenges of pharmacophore-based virtual screening are identifying a representative pharmacophore for an interaction and then identifying the compounds within a relevant chemical library that match the pharmacophore. ZINCPharmer is a pharmacophore search engine for purchasable chemical space that addresses both these challenges.

An interaction pharmacophore may be elucidated from a set of known active ligands by identifying a consensus pharmacophore that is conformationally accessible to all these ligands (1,4). These techniques do not require a ligand-bound structure, but may be computationally demanding if the input set contains many flexible

ligands. PharmaGist (5) is a free web server that can identify a consensus pharmacophore of a set of up to 32 ligands in a few minutes. Alternatively, structure-based approaches require a ligand-bound structure and identify a potential pharmacophore by analyzing the interaction site (6). ZINCPharmer provides a mechanism for deriving an initial pharmacophore hypothesis directly from structures within the PDB (Protein Data Bank), and also supports importing pharmacophore definitions developed using more computationally demanding approaches implemented in third-party tools.

Given a library of explicit compound conformations, conformers that match a 3D pharmacophore can be found using either fingerprint-based (7–9) or alignment-based (4,10) approaches. Fingerprints are well suited for similarity metrics (11), but, since they discretize the pharmacophore representation, provide inexact results. The EDULISS (12) online database provides fingerprint-based screening of a single-conformer library of a few million compounds, but the query fingerprint must be manually constructed from pairwise distance constraints. Alignment-based approaches produce more accurate and interpretable results, at the expense of more computation. For example, a library of fewer than a million conformers may take minutes or hours to screen (13). However, since there are substantially fewer protein targets than there are possible ligands, alignment-based pharmacophore screening can be used effectively when performing a reverse screen that identifies matching protein targets instead of ligands. PharmMapper (14) takes as input a single ligand and screens a database of over 7000 receptors for potential targets.

Both fingerprint and alignment-based approaches typically evaluate every conformer in the library, resulting in search times that scale with the size of the database. Newer methods, such as Pharmer (15) and Recore (16) use indexing approaches so that search times scale with the complexity and breadth of the query, not the size of the library. ZINCPharmer uses the open-source Pharmer software to enable the interactive search of more than 176 million conformations in just a few minutes, if not seconds.

\*To whom correspondence should be addressed. Tel: +1 412 383 5745; Fax: +1 412 648 3163; Email: dkoes@pitt.edu

## METHODS

ZINCPharmer searches a database of conformations calculated from the purchasable compounds of the ZINC database (17). ZINC is a comprehensive collection of commercially available, biologically relevant compounds suitable for screening. Purchasable compounds have an expected availability of <10 weeks and are either available from vendor stock or are make-on-demand. The ZINCPharmer library is synchronized with the ZINC library on a monthly basis. Compounds are both added and removed to maintain consistency and ensure that only currently purchasable compounds are retained. ZINC compounds are converted into 3D conformations using omega2 from OpenEye Scientific Software (<http://eyesopen.com>). Conformers are generated using the default settings and -rms.7, which improves the sampling of conformational space compared to the default setting of .5 (18). The 10 best conformers are saved.

The generated conformers are converted into an efficient search format using the Pharmer (15) open-source software. Pharmer identifies hydrophobic, hydrogen bond donor/acceptor, positive/negative ions and aromatic pharmacophore features using the SMARTS matching functionality of the OpenBabel toolkit (19). Currently,

the default set of SMARTS definitions is used, but these are subject to refinement based on user input. These features are stored in an efficient spatial index to support the rapid search of large chemical libraries. For example, the search shown in Figure 1 took less than 3 seconds.

The graphical user interface (Figure 1) for defining, refining and visualizing pharmacophore queries and their results is implemented using JavaScript and the Java-based Jmol (<http://www.jmol.org/>) molecular viewer. A modern, standards compliant browser with a recent Java plugin is required. Session state, which includes the pharmacophore definitions, can be saved in a human-readable JSON (JavaScript Object Notation) format and the aligned search results can be saved in the sdf molecular format. An internet forum hosts a user guide and provides technical support.

## DEFINING A PHARMACOPHORE QUERY

Using the Pharmer software, ZINCPharmer can automatically extract a set of pharmacophore features from molecular structure. Each feature consists of the feature type (hydrophobic, hydrogen bond donor/acceptor, positive/negative ion or aromatic), a position, and a search radius. Figure 2 illustrates the various methods

The screenshot shows the ZINCPharmer web application. At the top, there's a header bar with the URL <http://zincpharmer.csb.pitt.edu/pharmer.html?SESSION=http://zincpharmer.csb.pitt.edu/exam>. Below the header is a 3D molecular viewer showing a ligand (orange/red spheres) interacting with a receptor (blue/purple surface). In the bottom left, there's a pharmacophore query editor table:

Pharmacophore Class	x	y	z	Radius	Enabled
> HydrogenAcceptor	21.35	-14.53	19.63	1.00	<input checked="" type="checkbox"/>
> HydrogenAcceptor	19.36	-18.32	23.99	1.00	<input checked="" type="checkbox"/>
> HydrogenDonor	20.98	-16.95	18.75	1.00	<input checked="" type="checkbox"/>
> Negativelon	21.67	-15.08	20.61	1.50	<input checked="" type="checkbox"/>
> Negativelon	19.98	-19.40	22.84	2.00	<input checked="" type="checkbox"/>

To the right of the viewer is a results table titled "Results" showing 15 rows of data:

Name	RMSD	Mass	RBnds
ZINC19632959	0.589	196	7
ZINC01603015	0.598	176	7
ZINC00902211	0.636	181	7
ZINC19817592	0.644	188	6
ZINC01532722	0.653	190	8
ZINC04692839	0.654	148	4
ZINC12360700	0.657	190	4
ZINC20112430	0.664	196	7
ZINC32629480	0.666	158	4
ZINC02379324	0.667	182	5
ZINC03888065	0.667	195	4
ZINC02169444	0.670	164	5
ZINC03200680	0.671	174	7
ZINC03869280	0.672	183	5
ZINC00896091	0.680	160	6

At the bottom, there are buttons for "Add Feature", "Load Features...", "Load Receptor...", "Clear", "Load Session...", "Save Session...", and "Save Results...". A note at the bottom right says: "Please keep in mind that you may not redistribute major portions of ZINC without the express permission of John Irwin."

**Figure 1.** The ZINCPharmer interface. The Jmol-based molecular viewer is in the upper left and displays the pharmacophore features as spheres within the context of the interaction structure. A negative ion feature is shown in red mesh and the selected hydrogen acceptor in solid orange. Both a receptor structure, shown as a translucent partial-charge mapped surface, and a ligand structure, from which an interaction pharmacophore is automatically derived, may be uploaded. The pharmacophore query editor is shown in the bottom left and supports the interactive modification of the properties of the pharmacophore, including directions of hydrogen bonds and the size of hydrophobic regions. The full query session state can be saved and restored. Additional property filters, such as molecular weight, may be specified under the Filters tab while the visual styles of the molecular viewer may be set under the Viewer tab. The results browser is on the right and displays the ZINC id, which links directly to the ZINC database and purchasing information, the minimal RMSD of the compound pose to the query, the molecular weight and the number of rotatable bonds. The results may be sorted by any of the numerical features and the full set of result structures may be downloaded.

for creating an initial query. Features may be derived from a single ligand structure, a protein–ligand structure, a protein–protein structure or from the output of third-party software.

### From ligand structure

Any single-conformer molecular structure file that is compatible with OpenBabel (19) may be uploaded to define a set of pharmacophore features. All identified features of the molecule are enabled as pharmacophore query features. However, since by itself a ligand provides no information about the nature of an interaction, the result is not a true pharmacophore. For instance, even though low-energy conformers are often close in configuration to the bound structures (18), without additional information it is impossible to separate interacting features from non-interacting features. Instead, the pharmacophore derived from a single ligand structure should be thought of as a 3D similarity search.

If the receptor structure is known, a flexible docking of the ligand can generate a custom protein–ligand structure from which ZINCPharmer can automatically derive an interaction pharmacophore. Alternatively, if there are many known binders then a consensus pharmacophore can be elucidated (1,4) using software such as Chemical Computing Group's MOE (<http://www.chemcomp.com/>), Inte:Ligand's LigandScout (<http://www.inteligand.com/>), or PharmaGist (5) and the result can be imported into ZINCPharmer.

### From protein–ligand structure

When provided with both a receptor and bound-ligand structure, ZINCPharmer will automatically identify an

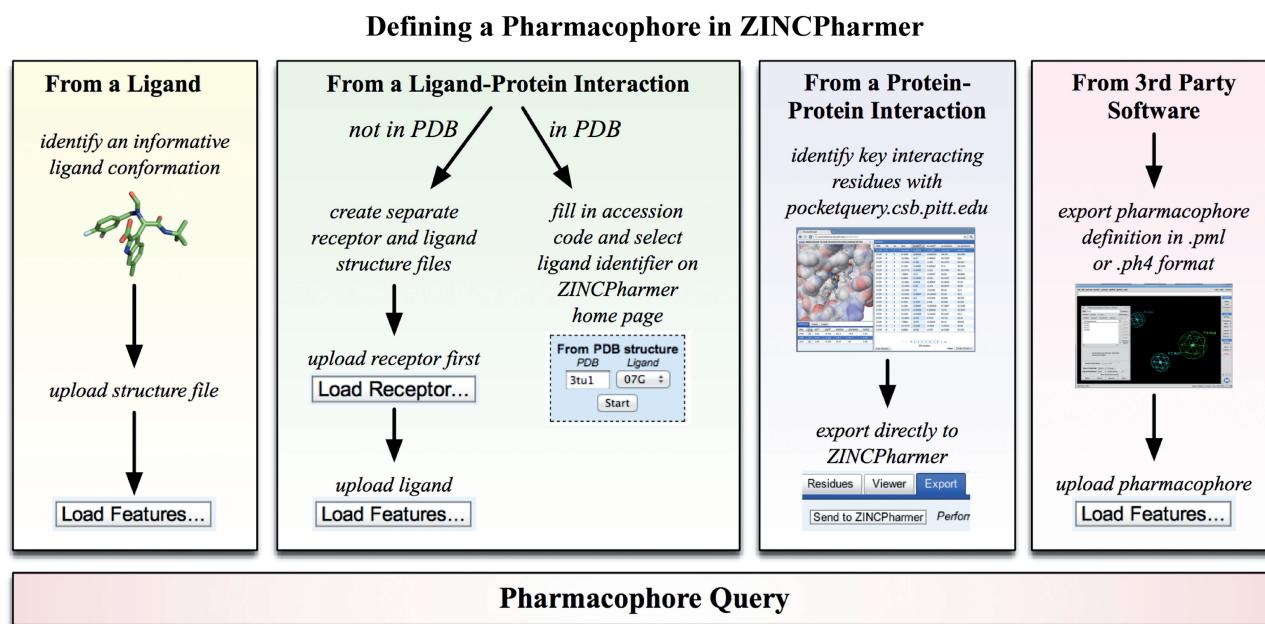
interaction pharmacophore. All possible pharmacophore features on the ligand are computed, but only those that are within a distance cutoff of complimentary features on the receptor are enabled. Hydrogen bond acceptors/donors must be within 4 Å of a hydrogen bond donor/acceptor on the receptor. Charged features must be within 5 Å of an oppositely charged feature on the receptor. Aromatic feature must be within 5 Å of a receptor aromatic feature. A ligand hydrophobic feature must be within 6 Å of at least three hydrophobic features on the receptor in order to require some degree of buriedness. The distance cutoffs are intended to be permissive and no angular cutoffs are applied since it is conceptually easier for a user to reduce the number of features in a pharmacophore query than to increase them (which requires investigating a much larger number of potential features).

If the protein–ligand structure exists in the PDB, then a shortcut is available on the ZINCPharmer home page where the user need only enter the PDB accession code, select the desired ligand and click the Start button (Figure 2). The corresponding ligand and receptor structures as well as their interaction pharmacophore will automatically be loaded into a new ZINCPharmer session.

For custom protein–ligand structures, for example, the result of a docking study, the receptor and ligand must be uploaded separately. In order to identify the interaction pharmacophore, the receptor must be uploaded first.

### From protein–protein interaction structure

ZINCPharmer is integrated with PocketQuery (<http://pocketquery.csb.pitt.edu>), a website that identifies protein–protein interaction (PPI) inhibitor starting points from PPI structure. Using a consensus scoring



**Figure 2.** Defining a pharmacophore query in ZINCPharmer. The Load Features button can be used to calculate the pharmacophore features of a ligand structure or to upload 3rd party pharmacophore definitions. Alternatively, an interaction pharmacophore can be derived directly from a ligand-bound structure in the PDB, or the essential pharmacophore of a protein–protein interaction can be exported from PocketQuery.

scheme (20), PocketQuery identifies a small set of interacting residues in a PPI structure whose mimicry by a small molecule is likely to inhibit the interaction. Within the PocketQuery interface, as shown in Figure 2, the selected set of residues can be exported directly to ZINCPharmer. The interaction pharmacophore between these ligand residues and the receptor will then be automatically generated as with a protein–ligand structure.

### From 3rd party software

ZINCPharmer includes support for uploading pharmacophore definitions represented in either PH4 format, used by MOE, or PML format, used by LigandScout. Additionally, the specialized mol2 format exported by PharmaGist (5) is recognized as a hybrid pharmacophore definition and ligand structure file. These programs can be used to elucidate a consensus pharmacophore from a set of active compounds. ZINCPharmer can then import the result and quickly identify all matching hits. However, there are several differences between the pharmacophore recognition routines and alignment policies of different software packages (21). In particular, the identification and positioning of hydrophobic features has the most variation between software packages. Consequently, ZINCPharmer searches using an externally defined pharmacophore will result in an overlapping, but not identical, set of hits compared with a search performed using the software that generated the pharmacophore.

## REFINING A QUERY

Although ZINCPharmer is capable of automatically extracting a pharmacophore from an interaction, it is expected that the user will further refine the query to enhance its specificity and applicability. This can be done by editing the properties of the query or by applying filters to the results.

### Query editor

Every pharmacophore feature is a row in the query editor and has a pharmacophore class (hydrophobic, hydrogen bond donor/acceptor, positive/negative ion or aromatic), a position specified in Cartesian coordinates, a radius representing the tolerance sphere to search around this position and an enabled/disabled setting. The pharmacophore query editor, shown in the bottom left of Figure 1, supports the interactive editing of these features, which are shown as spheres in the molecular viewer as seen in the top left of Figure 1. Features may be selected either in the query editor table or directly in the molecular viewer by clicking on the relevant sphere. Selected features may be batch processed (enabled, disabled, deleted or duplicated) through a contextual menu accessible by right-clicking the selected rows.

Some features have additional options unique to their pharmacophore class that are accessible through a drop down menu. Hydrogen bond donors/acceptors have an optional directionality, as shown in the drop down menu of Figure 1. The query vector is matched against a precomputed vector on the ligand. Since the actual

direction of the hydrogen bond is specific to the geometry of the interface, this match is necessarily approximate, and therefore a large tolerance in angular deviation is implemented by default.

Aromatic features also have an optional directionality constraint that matches against the normal vector of the aromatic ring. Hydrophobic features have an optional constraint for specifying the number of atoms participating in the hydrophobic area. For example, if a small hydrophobe, such as a methyl group, is desired, then the maximum number of atoms can be constrained to one. Alternatively, if a large, space-filling group is desired, such as an aliphatic ring, the minimum number of atoms can be constrained to five or higher.

### Filters

The results can be filtered both in terms of the number of returned results and the properties of the returned results. The number of hits can be reduced by specifying a limit on the number of different orientations returned for each conformation ('Max Hits per Conf'), the number of different orientations of different conformations returned for each molecule ('Max Hits per Mol'), or the total number of hits returned ('Max Total Hits'). In all cases, the search is terminated as soon as the limit is reached with no guarantee that the returned hits have the best possible root mean squared deviation (RMSD) to the query.

Each orientation of a conformer results from a different mapping and alignment of pharmacophore features on the ligand to the query features. If the query has many degrees of symmetry or tightly spaced features, reducing the number of orientations returned may substantially reduce the number of hits that need to be analyzed without omitting significant positional differences. Reducing the number of hits per a molecule is particularly useful when only the 2D properties of the results will be analyzed and only a single representative of each molecule is needed. Reducing the total number of hits is beneficial when the post-screening analysis is computationally intensive and only a sampling of the results is needed.

The results list can also be filtered by maximum RMSD. The orientation of the hits is computed using a weighted RMSD calculation (15), but the reported value is the unweighted RMSD between the calculated orientation and the query. Filtering by RMSD restricts the hits to those that have the best overall geometric match to the query. Additionally, hits can be filtered by the molecular properties of molecular weight (in Daltons) and number of rotatable bonds, both of which have been implicated as useful properties for identifying 'drug-like' molecules (22).

## PHARMACOPHORE SEARCH

Having defined a pharmacophore, searching for matching purchasable compounds is as simple as clicking the 'Submit Query' button. Searches take anywhere from a few seconds to a few minutes. Queries with more features, queries with many hydrophobic features (which are the most common features), queries with large search tolerances and symmetric queries (which require the

processing of many orientations per a matching conformer) will have longer search times. Results are returned and displayed in the results browser as they are found. An orientation of a conformer is only returned as a hit if all the matching features are within the specified search tolerances of the query when the conformer is aligned to minimize the weighted RMSD.

## RESULTS VISUALIZATION

The results of a search are displayed in the results browser shown in Figure 1. Each hit represents a unique orientation of a conformation to the query. For each hit, the ZINC identifier, RMSD to the query, molecular weight ('Mass'), and number of rotatable bonds ('RBnds') is shown. The ZINC identifier is a hyperlink that points to the corresponding compound web page in the ZINC database where purchasing information may be found. The results may be sorted by any of the numerical properties by clicking on the property heading in the results table. The complete set of oriented hits may be saved to an sdf file through the 'Save Results' button. The hits in this file are unordered and include the RMSD to the query as extra data attached to each molecule. This file is immediately useful as input to a secondary screening protocol such as ranking by energy minimization.

Individual hits are visualized with the query and a receptor (if present) by clicking on the corresponding row in the results browser. The viewer tab contains a wide assortment of colors and styles (wireframe, stick, spheres, etc.) for visualizing the results, the query ligand, the receptor residues and the receptor surface.

## DISCUSSION

The goal of ZINCPPharmer is to remove barriers to computational drug discovery. There is no need for users to purchase, install or build software: all that is needed is a modern web browser. Additionally, ZINCPPharmer takes care of the generation and storage of a large multi-conformer database of the biologically relevant and commercially available compounds of the ZINC database. Perhaps more importantly, the search performance of ZINCPPharmer (most searches take a few minutes, if not seconds) enables the iterative refinement of a pharmacophore hypothesis in the context of the entire chemical library. For example, users can quickly enable or disable features, adjust search tolerances and apply filters based on the results of previous searches to achieve a set of result compounds that has the desired size, specificity and chemical diversity. This sort of iterative refinement simply is not practical when searches take several hours. ZINCPPharmer enables a hands-on, experimental approach to developing a high-quality pharmacophore hypothesis that fully leverages the expertise and insight of the user. The matching compounds of the user-specified pharmacophore can then be purchased and experimentally validated as part of a broader drug discovery effort. ZINCPPharmer is a fully open access resource and is available at <http://zincpharmer.csb.pitt.edu>.

## ACKNOWLEDGEMENTS

The ZINC database is incorporated into ZINCPPharmer by permission. We thank NIH for supporting this resource.

## FUNDING

National Institutes of Health [GM097082] and National Institutes of Health [GM71896 to Brian Shoichet and John Irwin, supporting the ZINC database]. Funding for open access charge: National Institutes of Health [R01GM097082].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Leach,A.R., Gillet,V.J., Lewis,R.A. and Taylor,R. (2009) Three-dimensional pharmacophore methods in drug discovery. *J. Med. Chem.*, **53**, 539–558.
2. Mason,J.S., Good,A.C. and Martin,E.J. (2001) 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.*, **7**, 567–597.
3. Langer,T. and Krobat,E.M. (2003) Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Curr. Opin. Drug Discov. Dev.*, **6**, 370–376.
4. Wolber,G., Seidel,T., Bendix,F. and Langer,T. (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today*, **13**, 23–29.
5. Schneidman-Duhovny,D., Dror,O., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2008) PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Res.*, **36**, W223–W228.
6. Wolber,G. and Langer,T. (2004) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.*, **45**, 160–169.
7. Mason,J.S. and Cheney,D.L. (2000) Library design and virtual screening using multiple 4 point pharmacophore fingerprints. In *Pacific Symposium on Biocomputing*, **5**, 576–587.
8. Baroni,M., Cruciani,G., Scialoba,S., Perruccio,F. and Mason,J.S. (2007) A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.*, **47**, 279–294.
9. Floris,M., Masciocchi,J., Fanton,M. and Moro,S. (2011) Swimming into peptidomimetic chemical space using pepMMsMIMIC. *Nucleic Acids Res.*, 10.1093/nar/gkr287.
10. Raymond,J.W. and Willett,P. (2002) Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput. Aided Mol. Des.*, **16**, 521–533.
11. Sheridan,R.P. and Kearsley,S.K. (2002) Why do we need so many chemical similarity search methods? *Drug Discov. Today*, **7**, 903–911.
12. Hsin,K.Y., Morgan,H.P., Shave,S.R., Hinton,A.C., Taylor,P. and Walkinshaw,M.D. (2011) EDULISS: a small-molecule database with data-mining and pharmacophore searching capabilities. *Nucleic Acids Res.*, **39**, D1042–D1048.
13. Fang,X. and Wang,S. (2002) A web-based 3D-database pharmacophore searching tool for drug discovery. *J. Chem. Inf. Comput. Sci.*, **42**, 192–198.
14. Liu,X., Ouyang,S., Yu,B., Liu,Y., Huang,K., Gong,J., Zheng,S., Li,Z., Li,H. and Jiang,H. (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.*, **38**, W609–W614.
15. Koes,D.R. and Camacho,C.J. (2011) Pharmer: efficient and exact pharmacophore search. *J. Chem. Inf. Model.*, **51**, 1307–1314.
16. Maass,P., Schulz-Gasch,T., Stahl,M. and Rarey,M. (2007) Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. *J. Chem. Inf. Model.*, **47**, 390–399.

17. Irwin,J.J. and Shoichet,B.K. (2005) ZINC- a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, **45**, 177–182.
18. Kirchmair,J., Wolber,G., Laggner,C. and Langer,T. (2006) Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.*, **46**, 1848–1861.
19. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminforma.*, **3**, 33.
20. Koes,D.R. and Camacho,C.J. (2012) Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure. *Bioinformatics*, **28**, 784–791.
21. Spitzer,G.M., Heiss,M., Mangold,M., Markt,P., Kirchmair,J., Wolber,G. and Liedl,K.R. (2010) One concept, three implementations of 3D pharmacophore-based virtual screening: distinct coverage of chemical search space. *J. Chem. Inf. Model.*, **50**, 1241–1247.
22. Veber,D., Johnson,S., Cheng,H., Smith,B., Ward,K. and Kopple,K. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.*, **45**, 2615–2623.